

指向性を持って化合物の特性を進化させる 仮想ライブラリの構築システムとその web サービス化

新井 直樹^{†1,a)} 吉川 舜亮^{†2} 安尾 信明^{†2}
中嶋 悠介^{†2} 吉野 龍之介^{†3} 関嶋 政和^{†2,3,b)}

概要: 創薬研究においては標的蛋白質に対するヒット化合物を探索した後、リード最適化などによってその物性を薬物らしく変化させていく。しかしその際には理論上膨大な量の化合物を含む化合物ライブラリが必要となる上に、その天文学的数の化合物を実際に合成実験することや、そのライブラリを計算機で処理することはほぼ不可能である。そこで本研究では、実際に有機合成可能な化学反応をもとにした仮想的な反応ルールを入力化合物に対して多段階に適用することで、入力化合物の物性を変化させた新規周辺化合物を生成し、そのライブラリを構築するシステムを開発した。また、ライブラリ構築の際に用いられる反応ルールに一定の傾向を与えることによって効率的に目的とする物性を向上させた周辺化合物ライブラリの構築が可能となる。更に、このシステムを web サービスとして公開することで、ユーザがヒット化合物を得た後に煩雑な操作なしにライブラリ構築や標的蛋白質に対し特異的に阻害脳を示す構造の保護、目的とする化合物の合成経路の探索を行うことを可能とする。

キーワード: 創薬, 仮想化合物ライブラリ, web サーバー

System to evolve properties of ligands and compile virtual libraries and implementation it as web services

NAOKI ARAI^{†1,a)} SHUNSUKE YOSHIKAWA^{†2}
NOBUAKI YASUO^{†2} YUSUKE NAKASHIMA^{†2}
RYUNOSUKE YOSHINO^{†3} MASAKAZU SEKIJIMA^{†2,3,b)}

Abstract: In this study, we construct system to evolve properties of ligands by applying synthetic rules involved in the improving Hydrogen Bonding Donor (HBD), Hydrogen Bonding Acceptor (HBA) and stability in multiple stages, thereby compiling virtual compound libraries. In addition, we implement it as web services. After the user obtain hit compounds, our services make it possible for user to obtain virtual libraries of drug-like compounds and check those property and applied virtual chemical reactions interactively. In this paper, we explain system method and operating procedure of web service.

Keywords: Drug discovery, Virtual compound library, web server

1. はじめに

創薬研究においては、まず標的蛋白質に対してヒット化合物の探索が行われるが、探索の結果得たヒット化合物は必ずしも薬品としてふさわしい活性や物性を十分に持っているとは限らない。そのため、ヒット化合物を獲得した後にリード最適化や構造活性相関 (SAR) の解析などによって活性や物性を改善することで、ヒット化合物をドラッグライクに進化させたリード化合物を獲得する。

しかし、理論的な化合物数は約 10^{60} 個と天文学的な数となるため[1], 現在報告されている大規模な仮想化合物ライブラリにおいても 10^{20} 個とその数に遠く及ばない[2]。また、

もしその必要数を満たすライブラリが得られたとしても、 10^{60} 個にも及ぶ化合物を全て合成して実験を行うことは不可能であり、その膨大なデータを利用したドッキングなどのシミュレーションを行うことも難しい。

そこで本研究では、実際に有機合成可能な化学反応を基にした仮想的な化学反応をヒット化合物に対して適用することによって、ヒット化合物をドラッグライクに進化させた周辺化合物を獲得しそのライブラリを構築すること、またその際にヒット化合物に対し一定の傾向を持った反応を適用することによって、効率的に目的の物性が向上した周辺化合物を得ることを目的とする。

さらに、本研究で開発したシステムを web サービスとして実装することで、ユーザがヒット化合物を獲得した後に、インタラクティブに化合物の物性値や合成ルールを確認しながらヒット化合物から合成可能なドラッグライク化合物群を獲得可能にすることも目的としている。

2. 手法

本研究では、図 1 のように、仮想的な化学反応を入力化

†1 東京工業大学 工学部 情報工学科
Department of Computer Science, School of Engineering, Tokyo Institute of Technology

†2 東京工業大学 大学院情報理工学専攻 計算工学専攻
Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

†3 東京工業大学 学術国際情報センター
Global Scientific Information and Computing Center, Tokyo Institute of Technology

a) arai.n.ai@m.titech.ac.jp

b) sekijuma@gsic.titech.ac.jp

化合物に対して多段階に適用していくことで新たな周辺化合物を獲得し、ライブラリを成長させていく[3]。さらに、その際に特定の物性を向上させるような傾向を持った反応ルールを重点的に適用することで、入力された化合物の物性を目的とする方向に進化させる。

本システムは Python 言語で記述されている。化合物構築の際に適用する反応ルールは実際に有機合成可能な化学反応を元として構築した仮想的な反応を SMARTS 記法[4]で記述し、それをケモインフォマティクスのツールキットである RDKit[5]によって入力化合物に適用する。また、web システムの構築の際には web フレームワークとして Django[6]を利用した。

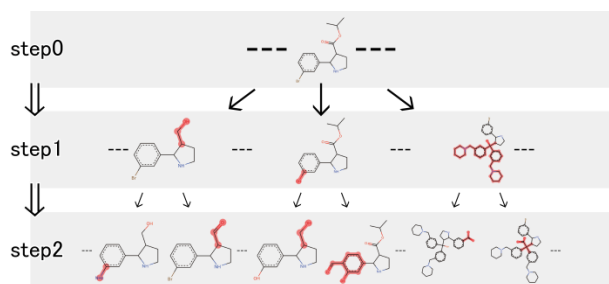


図 1 ライブラリ構築の模式図

3. ライブラリ構築の実験

3.1 実験条件

システムによる化合物構築検証を行うため、入力化合物として、ナミキ商事の building block 統合データベース (2014 年 10 月版) に含まれる約 480000 個の化合物を均等に三分割したライブラリから規則正しく 160 個飛ばしで抽出した、1000 化合物のライブラリ 3 種類を作成した。

また、適用する反応として、仮想化学反応 193 種類のうち全てのルール、Hydrogen Bonding Donor (HBD) 増加傾向の 31 種類、Hydrogen Bonding Acceptor (HBA) 増加傾向の 163 種類、環構造増加傾向の 22 種類、の 4 つのルールセットを利用する。

3.2 計算条件

CPU : Intel Core i7-3770 (3.40GHz * 4)

メモリ : 16.8GB

言語 : Python, RDKit

OS : Ubuntu 15.04 (64 bit)

3.3 入力ライブラリごとの比較

前述の 3 セットの入力に対して 193 種類全ての反応を適用した場合の実行時間および得られる化合物数を比較した表 1 を見ると、#0 のライブラリを入力として利用した場合に、他のライブラリを入力した場合と比べて得られる化合物数及び実行時間が増大していることが分かる。これは、出力化合物数や実行時間が入力された化合物への反応適用可能回数に依存することが原因となっている。

ここで各入力ライブラリについて、出力化合物の反応適

用段階ごとの物性の平均値を図 2 のように比較すると、すべてのライブラリにおいて概ね同様の傾向で物性値が変化していることが分かる。

表 1 入力ライブラリごとの出力化合物数と実行時間

入力ライブラリ	出力化合物数[#]	実行時間[sec]
#0	2620659	40250.35
#1	1633059	7169.08
#2	1799410	8503.77

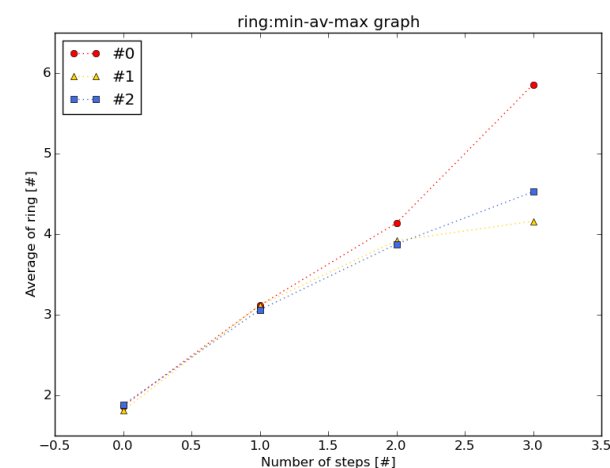
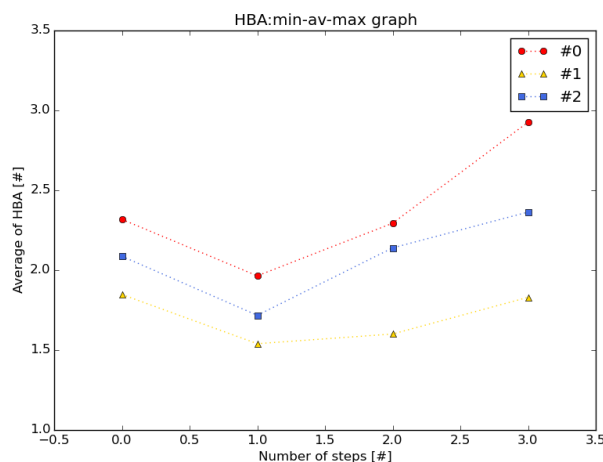
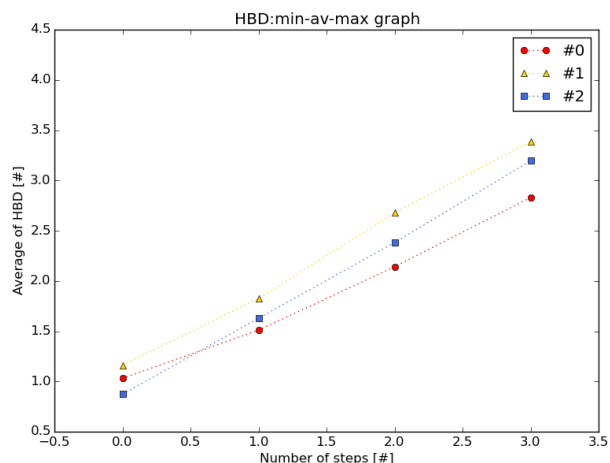


図 2 入力化合物ごとの出力化合物の平均物性値 (HBD, HBA, 環構造数) の比較

以上から、入力されたライブラリごとに出力化合物数や反応を適用した際の平均的な物性値の変化量に差があるものの、物性値変化の傾向は概ね同様であるため、次に行う適用するルールセットごとの比較においては3種類の入力化合物セットのうち#0 を利用した場合についてのみ述べる。

3.4 適用するルールごとの比較

それぞれの反応セットを適用した場合の実行時間と化合物数を比較したものが表2である。この表から、特定の物性日課した傾向を持った反応セットを適用することによって実行時間が0.026~0.43倍に減少し、化合物数も0.011~0.58倍に減少していることが分かる。

これは入力化合物に対して多段階に選択された反応ルールセットを適用して新規化合物を獲得しているために、利用する反応ルールを目的とする物性に特化したものに絞り、適用する反応ルールが減少したことが原因となっている。

しかし、図3のようにそれぞれの反応セットが目的とする物性について出力化合物のライブラリ内での存在比率を比較すると、それぞれ目的とする物性の多い化合物を獲得可能であり、また目的とする物性の少ない化合物を減少させていることがわかる。

表 2 特定の物性に特化した反応セットを適用した場合の出力化合物数と実行時間

反応セット	出力化合物数[#]	実行時間[sec]
すべてのルール	2620659	40250.35
HBD 増加傾向	1512092	17144.35
HBA 増加傾向	28766	131.09
環構造増加傾向	47585	104.88

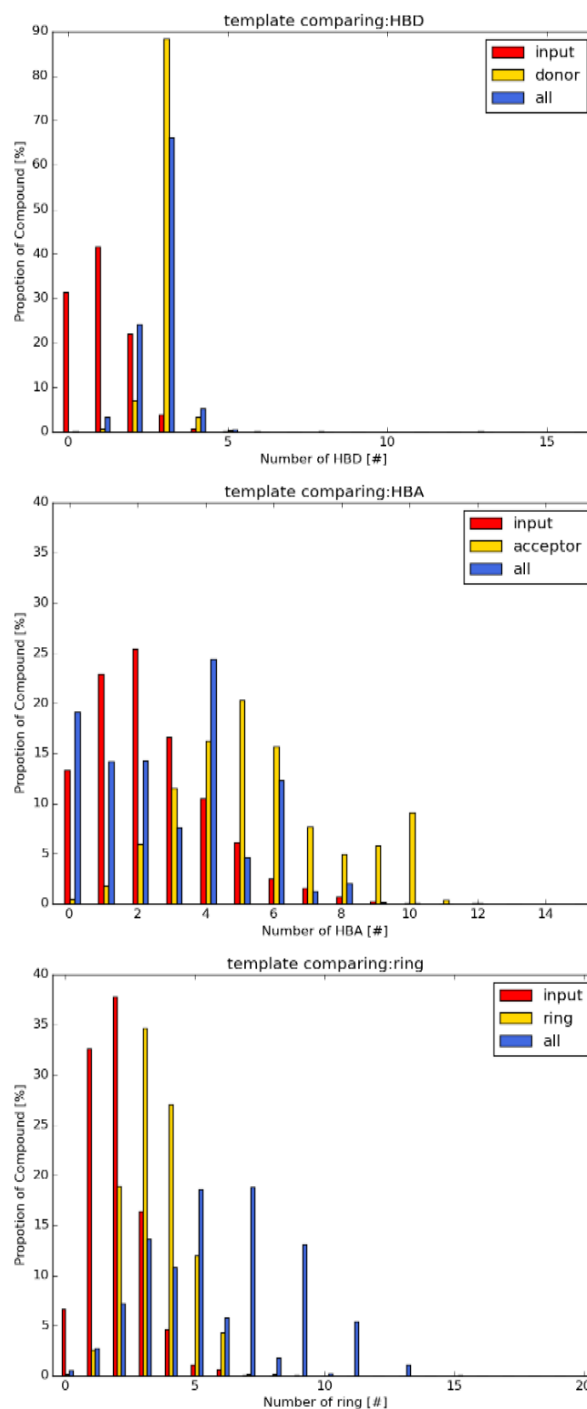


図 3 特定の物性に特化した反応セット (HBD, HBA, 環構造増加傾向) を適用した場合の目的とする物性を持つ出力化合物のライブラリ存在比率の比較

4. web インターフェース

本研究で構築したシステムを web サービスとして公開することで、ユーザがヒット化合物を獲得した後に hit-to-lead やリード最適化、SAR の解析に利用可能な仮想化合物ライブラリを構築可能にする。

サービス利用の際に煩雑な操作やアカウント作成などといったユーザ登録は必要なく、入力された化合物情報及び構築されたライブラリ等の情報は利用後自動的に削除される。また、必要であればライブラリ構築の詳細な設定を行うことも可能である。

4.1 化合物入力

ライブラリ構築の起点となる化合物を複数の方法 (SMILES, mol, mol2, sdf 及びそれらを含む zip, gzip ファイル) で入力する。この時入力された化合物データは分子量や logP などの物性と構造情報 (SMILES) に変換され、システムに一時保存される。

入力した化合物の物性や構造はその場で確認することが可能であり、実際のライブラリ構築の際に起点となる化合物として利用するかどうかの指定も可能である。

実際に表示される画面は図 4 のようになっている。

check	name	structure	active	logP	MW	HBD	HBA	ring	Rule of Five				
									P	W	D	A	#
<input type="checkbox"/>			<input type="radio"/>	1.924	201.22	2	1	1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	4
<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	1.564	203.6	2	1	1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	4
<input type="checkbox"/>			<input type="radio"/>	0.833	157.19	2	1	1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	4
<input type="checkbox"/>			<input type="radio"/>	2.322	288.14	2	1	1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	4

図 4 化合物入力の表示画面

4.2 指定構造の保護

ヒット化合物と標的タンパク質との相互作用様式などから、特定の相互作用を保護したまま、新規の相互作用を獲得するような構造の変化を行うことを可能にするため、特定の官能基や母核となる構造に対して保護を行う。化合物に対して反応ルールを適用して得られた化合物が指定された保護構造を保持していない場合に、それを除去することで、構造が保護された新規化合物のみを得ることが可能になる。

保護構造の指定は通常の化合物入力と同様に行う。また、

高速化のために反応の段階ごとに構造の保護を行うか、多様性を得るために反応が全て終了してから構造が保護されていない化合物を排除するかが指定可能で、さらに入力された構造のうち保護する構造の個数も設定可能である。

ここまでの入力で 1 つの保護ライブラリが完成し、この保護ライブラリを複数作成することでライブラリ構築の際に複雑な構造保護を行うことも可能となっている。

実際に表示される画面は図 5 のようになっている。

check	structure	logP	MW	HBD	HBA	ring
<input type="checkbox"/>		1.748	84.14	0	0	1
<input type="checkbox"/>		-0.299	46.03	1	1	0

図 5 指定構造の保護画面

4.3 ライブラリ構築

全 193 種類の反応から、ライブラリ構築の際に適用する反応を選択する。全反応の中から自由に選択することも勿論可能であるが、事前に用意された、特定の物性の向上に特化した反応セットを利用することで効率的に目的とするライブラリを構築可能となる。

また、その反応をライブラリに適用する回数を指定する。通常、三段階反応を適用するが、適用回数を増加させることで実行時間が増加するものの、より多くの化合物を獲得可能になる。

適用する反応とその適用段階数を設定したあとは、ライブラリ構築開始ボタンを押すだけで簡単にライブラリ構築が開始する。

実際に表示される画面は図 6 のようになる。

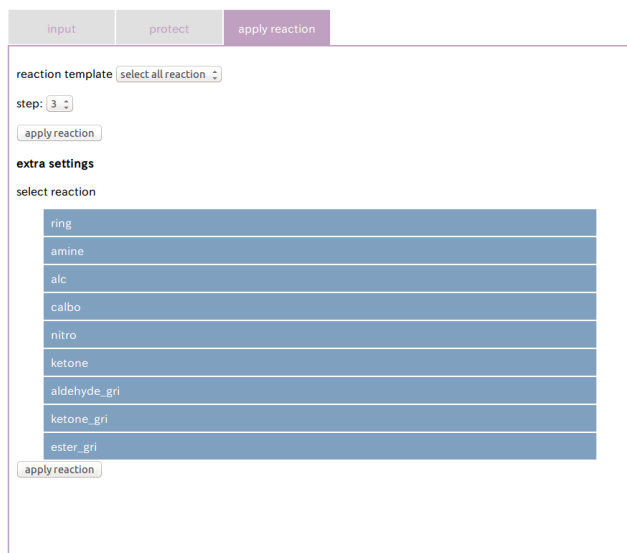


図 6 ライブラリ構築画面

4.4 ライブラリに含まれる化合物の表示

構築されたライブラリに含まれる化合物を一覧にして表示する。図 7 のように、化合物の名称や構築における役割、化合物が生成された段階数、また RDKit の組み込み関数を利用して生成されたモル分子量・logP・HBD・HBA・環構造の数・Lipinski's Rule of 5 のプロパティを表示する。化合物の名称は、その化合物が初めて生成されたときの入力化合物の名称またはライブラリ内での番号に、適用された反応の番号を付記したものとなっている。

一度に表示できる化合物数は 20 個までとなっているが、指定した番号の化合物にジャンプする機能も実装されており、また 'trace evolution' のリンクをクリックすることによって次項で述べる進化経路の表示機能に遷移することが可能である。

compound index:6-25													
name link	structure	status	step	logP	MW	HBD	HBA	ring	Rule of Five				
		is	is						M	D	A	S	
trace evolution #6		(X)	(0)	0.5	185.17	2	1	1	(0)	(0)	(0)	(0)	4
trace evolution #7		(0)	(0)	1.924	201.22	2	1	1	(0)	(0)	(0)	(0)	4
trace evolution #8		(0)	(X)	1.564	203.6	2	1	1	(0)	(0)	(0)	(0)	4
trace evolution #9		(0)	(0)	0.033	157.19	2	1	1	(0)	(0)	(0)	(0)	4
trace evolution #10		(0)	(0)	2.322	288.14	2	1	1	(0)	(0)	(0)	(0)	4
trace evolution #11		(X)	(X)	1	0.274	265.27	2	3	2	(0)	(0)	(0)	4
trace evolution #12		(X)	(X)	1	-0.09	270.28	3	1	1	(0)	(0)	(0)	4
trace evolution #13		(X)	(X)	1	0.704	227.26	2	0	1	(0)	(0)	(0)	4
trace evolution													

図 7 ライブラリに含まれる化合物の表示画面

4.5 反応過程の表示

得られたライブラリについて解析を行い目的とする化合物群が得られた後に、その化合物を実際に合成しその標的蛋白質への阻害能を検定することで入力化合物からリード最適化された化合物が入手できる。

そこで、本サービスでは図 5 のようにある化合物の反応適用前後の化合物を、その生成に利用した反応やその反応による変化部位とともに表示する。この機能を利用することによって、入力化合物から目的の化合物への合成経路を探索し、実際に目的の化合物を有機合成することが可能となる。

また、表示された反応適用前後の化合物それぞれについて、'trace evolution' のリンクや 'view property' のリンクが用意されており、それぞれクリックすることでその化合物の進化経路の表示や物性値等の表示が可能である。

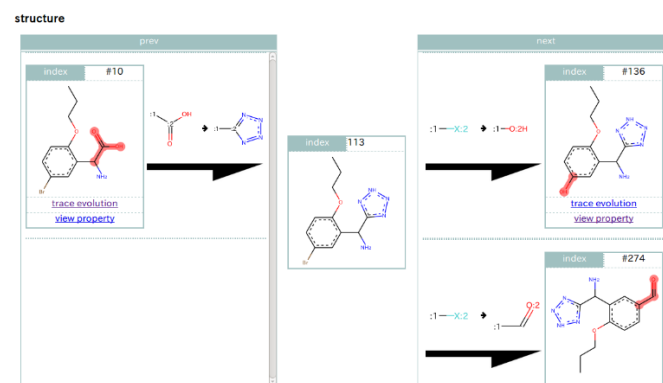


図 8 化合物進化過程の表示画面

4.6 ライブラリのダウンロード

構築した仮想化合物ライブラリを複数の形式 (SMILES, mol, sdf) でダウンロードすることが可能であり、また、構築したライブラリだけでなく、入力化合物のライブラリや化合物ライブラリの物性一覧のダウンロードも可能である。

このサービスで構築した仮想化合物ライブラリをドッキング処理や構造活性相関の解析などを行う他の創薬支援ソフトウェアで利用することによって、より効率的な創薬研究が可能になると考えられる。

5. おわりに

本研究で構築されたシステムは、実際に有機合成可能な反応を基にした仮想的な化学反応を入力された化合物に対して多段階に適用することによって新たな化合物を得、そのライブラリを構築する。また、この際に行う化学反応に特定の物性に特化した傾向を持たせることによって、効率よく目的とする物性を持った化合物を多く含んだライブラリを構築可能である。

更にシステムを web サービスとして実装することで、リード最適化や構造活性相関の解析を補助する仮想化合物ライブラリを煩雑な操作なしに構築することを可能とした。

また、ライブラリの構築だけでなく、構築段階における

特定構造の保護、ライブラリ内化合物の閲覧、化合物構築経路の追跡など、創薬研究において有用な機能が利用可能である。

しかし、化合物描画ソフトなどを用いた構造入力やライブラリ構築段階での物性値を参照したフィルタリング機能、物性値によるランキングなど、本サービスのインターフェースや機能には改良・追加すべき点が多く存在している。

また、本サービスでは 193 種類の反応ルールが実装されているものの、全ての反応ルールを利用した場合でも周辺化合物を完全に網羅することは不可能であり、更に適用された反応を再現するために必要な試薬や先行研究による合成反応の使用例や反応ルールの適用箇所ごとの合成可能性の検討など、既に実装されている反応の情報量も不足していて、実際に有機合成を行うに当たって重要な情報を網羅できていない。

よって、既存の反応データベースや有機合成に関する文献などを参照して反応に関する情報や反応の数を豊かにしていくことは勿論のこと、ユーザからの要望に応じていく形でサービスのインターフェースや機能を拡充していくことが重要な今後の課題となると考えている。

参考文献

- [1] Regine S. Bohacek, Colin McMartin, and Wayne C. Guida: The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective, Medicinal Research Reviews, Vol.16, No.1, pp.3-50, 1996.
- [2] Zhengwei Peng: Very large virtual compound spaces: construction, storage and utility in drug discovery, Drug Discovery Today: Technologies, Vol.10, No.3, pp.387-394, 2013.
- [3] 吉川舜亮, 安尾信明, 吉野龍ノ介, 関嶋政和: 方向性を持たせたグラフ構造変換による仮想化合物ライブラリの構築の研究, 研究報告バイオ情報学 (BIO), Vol.2015-BIO-41, No.12, pp.1-6, 2015-03-13.
- [4] Daylight Theory: SMARTS - A Language for Describing Molecular Patterns, <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>
- [5] RDKit: Open-Source Cheminformatics Software, <http://www.rdkit.org/>
- [6] The Web framework for perfectionists with deadlines | Django, <https://www.djangoproject.com/>